Synthetic Biology, 2019, 4(1): ysz025

doi: 10.1093/synbio/ysz025
Advance Access Publication Date: 29 October 2019
Research article

Highly multiplexed, fast and accurate nanopore sequencing for verification of synthetic DNA constructs and sequence libraries

Andrew Currin [1,2,†], Neil Swainston [1,2,3,†], Mark S. Dunstan[1,2], Adrian J. Jervis [1,2], Paul Mulherin [1,2], Christopher J. Robinson [1,2], Sandra Taylor[1,2], Pablo Carbonell [1,2], Katherine A. Hollywood [1,2], Cunyu Yan[1,2], Eriko Takano [1,2], Nigel S. Scrutton [1,2], and Rainer Breitling [1,2,*]

[1]Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, The University of Manchester, Manchester M1 7DN, UK, [2]School of Natural Sciences, Department of Chemistry, Faculty of Science and Engineering, The University of Manchester, Manchester M13 9PL, UK and [3]Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK

*Corresponding author: E-mail: rainer.breitling@manchester.ac.uk

†These authors contributed equally to this work.

Abstract

Synthetic biology utilizes the Design–Build–Test–Learn pipeline for the engineering of biological systems. Typically, this requires the construction of specifically designed, large and complex DNA assemblies. The availability of cheap DNA synthesis and automation enables high-throughput assembly approaches, which generates a heavy demand for DNA sequencing to verify correctly assembled constructs. Next-generation sequencing is ideally positioned to perform this task, however with expensive hardware costs and bespoke data analysis requirements few laboratories utilize this technology in-house. Here a workflow for highly multiplexed sequencing is presented, capable of fast and accurate sequence verification of DNA assemblies using nanopore technology. A novel sample barcoding system using polymerase chain reaction is introduced, and sequencing data are analyzed through a bespoke analysis algorithm. Crucially, this algorithm overcomes the problem of high-error rate nanopore data (which typically prevents identification of single nucleotide variants) through statistical analysis of strand bias, permitting accurate sequence analysis with single-base resolution. As an example, 576 constructs (6 × 96 well plates) were processed in a single workflow in 72 h (from *Escherichia coli* colonies to analyzed data). Given our procedure's low hardware costs and highly multiplexed capability, this provides cost-effective access to powerful DNA sequencing for any laboratory, with applications beyond synthetic biology including directed evolution, single nucleotide polymorphism analysis and gene synthesis.

Key words: synthetic biology; next-generation sequencing; DNA assembly; nanopore sequencing; strand bias.

Submitted: 8 July 2019; **Received (in revised form):** 1 October 2019; **Accepted:** 3 October 2019

Introduction

Synthetic biology is the engineering of biological systems for a desired outcome. These outcomes can include a variety of applications, including the production of therapeutics (1), fine chemicals (2, 3), biofuels (4, 5) and biomaterials (6, 7) or the generation of novel functional organisms like biosensors (8) or computers (9). Typically, synthetic biology requires the construction of synthesized DNA into specifically designed constructs, which are often large and complex. The discipline is fueled by the availability of inexpensive DNA synthesis and assembly methods (optionally involving high-throughput automation), allowing any designed sequence to be synthesized and assembled in a matter of days.

There are a wide variety of DNA assembly techniques available to create synthetic DNA constructs, including Ligase Cycling Reaction (LCR), Golden Gate, BioBricks, Gibson assembly (and other similar methods based on joining homologous ends) and recombination (10–12). For all of these methods, transformation of a host (e.g. *Escherichia coli*) permits the screening of clones to identify those harboring the correct assembly. Following this, identified constructs must be analyzed to verify that their sequence is identical to the intended design prior to testing for functional performance. Ideally, the sequencing of these large constructs is performed in-house, to allow tight coupling to the construct assembly and phenotyping platforms. Next-generation sequencing (NGS) is ideally suited to this task, given its ability to rapidly analyze gigabases of sequence data (13, 14). To date, NGS has been utilized in synthetic biology for a range of applications, including the design and analysis of synthetic biology parts (15–17), the characterization of genetic parts and circuits (18) and the study of DNA logic functions and circuits (19–21).

A number of NGS technologies are available, capable of generating both short- (notably Illumina, Ion Torrent) and long-read data (Pacific Biosciences and Oxford Nanopore) (22, 23). Sequence data derived from nanopore technology generates the longest read lengths currently available (often over 100 kb), allowing for the easy identification of repeating or moveable elements, which is challenging for short-read data (24). This would perfectly meet the needs of synthetic biology, which due to design-of-experiment (DoE) and combinatorial assembly approaches constructs can contain a large number of repeated elements and sequences (25, 26). However, the high-error rate of nanopore data (currently at 5–15%, compared to $<$1% for short read technologies) means that it is currently unsuitable for accurate detection of single nucleotide variants (SNV) and indels (27), unless combined with accurate short-read data (termed 'hybrid assemblies' (28–32)). Unfortunately, the large costs associated with installing high-accuracy sequencing technologies (e.g. Illumina or PacBio) limits the in-house accessibility of rapid and accurate NGS for small- to medium-sized synthetic biology laboratories.

This work describes a standardized workflow for the accurate sequence verification of synthetic DNA assemblies using nanopore technology, that is both quick and low cost and offers a solution to the problems created by the inherent high-error rate. The process involves polymerase chain reaction (PCR) amplification of assemblies directly from *E. coli* transformants to quickly obtain purified amplicons, nanopore sequencing to generate real-time data and a novel analysis algorithm to validate correct assemblies with high accuracy. Specific DNA barcodes are designed for the workflow, enabling highly multiplexed sequencing of hundreds of large constructs. Importantly, this novel analysis can accurately discriminate between systematic sequencing errors (inherent in high-error rate nanopore data) and genuine assembly mutations, permitting accurate SNV identification from high-error rate nanopore data. The process is designed to sequence any construct made using BglBrick vectors (33), a standardized set of *E. coli* expression plasmids widely used for the assembly of constructs in synthetic biology. To our knowledge, this is the first NGS workflow specifically designed to meet the needs of high throughput, multi-fragment DNA assembly approaches.

Materials and methods

Design of multiplexing primers

Universal primer binding sequences that are common to all the BglBrick vectors were first identified, such that they were upstream and downstream from the multiple cloning sites (MCS) of every construct (Supplementary Materials S1 and S2). Consequently, the same sequences could be employed to amplify any insert cloned into any of the BglBrick vectors. Whilst a terminal 3′ G was not required for all vector templates; it was found that its inclusion in the primer sequence significantly improved primer performance and it was therefore included in all designs.

For barcode design, 500 random 24-nucleotide sequences were designed and added to the forward and reverse universal primer binding site sequences to create the complete primer sequences. Primers were ranked according to their propensity to form secondary structure (highest ΔG ranked first), according to DINAMelt calculations (34). From the best-ranked sequences, 96 were selected as reverse barcode primers. 6×96 primer pairs were then created, providing 576 unique primer combinations, which can all be combined together for sequencing on one NGS run (see Supplementary Files). Further primers have been identified so that more than 6 forward barcode primers can be used to extend the multiplexing capability (Supplementary Material S3).

Construct assembly

Specifically designed constructs were assembled using the LCR methodology as previously described (3, 35, 36). Transformant *E. coli* colonies were selected by automated colony picking into 1 ml Lysogeny Broth (with appropriate antibiotic added), covered with a breathable seal (Greiner) and incubated overnight at 30°C (950 rpm). Cultures were then diluted 1:400 in dilute phosphate buffered saline solution (13.7 mM NaCl, 1 mM sodium phosphate, 0.27 mM KCl, pH 7.4) to generate the PCR templates.

Culture PCR for sample barcoding

PCR conditions were optimized for robust amplification from diluted overnight cultures. The best performance was obtained using CloneAmp HiFi (Takara Clonetech) high-fidelity polymerase, prepared as a 2× premix. Reactions contained 5 µl enzyme premix, 2.5 µl primer mix (containing 2 µM of both forward and reverse primers) and 2.5 µl diluted PCR template. PCR was performed using a 96 well thermocycler (Eppendorf Mastercycler) with an initial denaturation incubation at 95°C for 180 s, followed by 35 cycles of 98°C for 20 s, 64°C for 15 s and 72°C for 210 s, concluding with a final incubation of 72°C for 210 s. Amplicons were then analyzed by capillary electrophoresis, using the Fragment Analyzer Automated CE System and the

dsDNA 930 reagent kit, following the manufacturer's instructions.

Sequencing

Amplicons selected for sequencing (entire plates or selected samples) were pooled and purified using the NucleoSpin Gel and PCR clean-up kit (Macherey Nagel). A total of 1–1.5 μg DNA was then prepared using the 1D amplicon/cDNA by ligation kit (SQK-LSK109, Oxford Nanopore) and sequenced using the MinION device (R9.4.1 flow cell) following the manufacturer's instructions. During the MinION run the generation of real-time data permits rapid processing and analysis of data for the fast identification of correct assemblies (37). Typically, data were collected for up to 24 h and basecalling of the raw data was performed using Guppy (v2.3.7). More data could be obtained by increasing sequencing time to 48 h.

Data analysis

A bespoke data analysis algorithm was constructed using both new and existing tools. The process is split into a number of steps: sequence reading, demultiplexing, alignment and analysis. Processed (basecalled) sequence data are initially read using Biopython (38). Demultiplexing is performed by custom code that is tolerant of the high-error rate typically found in nanopore data. Once demultiplexed, sequences are aligned to either their known target sequence (in the case of sequence verification) or all supplied target sequences (in the case of sequence identification) using the freely available Burrows-Wheeler Aligner program, BWA-MEM (39). Variant Call Format (vcf) files are then generated using SAMtools mpileup method (40), and the resulting vcf files are summarized to provide a measure of sequence identity, mutations, indels and maximum read depth for each of the individual samples within the pool.

The data analysis algorithm considers forward and reverse reads separately, due to the high frequency of strand bias in miscalling of bases. A Bayesian statistical strategy is introduced to ensure that SNVs are reported only if they appear concordantly on both strands. At positions of high-strand bias, indicating unreliable sequencing data, it is assumed that the nucleotide is most likely that of the template sequence. In the case of low strand bias and high data reliability, nucleotide calling probabilities are calculated, estimating the likelihood associated with each nucleotide for each read direction separately, using a binomial distribution, resulting in an implicit weighting by the number of reads on each strand. As a consequence, SNVs are only confidently reported with high probabilities if they are concordant on both strands and supported by an adequate number of reads. Source code and instructions are openly available at https://github.com/neilswainston/sbc-ngs/.

Results and discussion

Design of barcoded primers

The objective of this study was to develop an NGS workflow that could quickly and accurately determine host transformants exhibiting the correctly assembled synthetic DNA sequence. This process depends on PCR amplification of assemblies directly from the culture in order to quickly obtain concentrated DNA amplicons for sequencing, thus eliminating more laborious plasmid extraction protocols.

In order to sequence many discrete samples in a single experiment (multiplexing), NGS employs an indexing system based on DNA 'barcodes'. Typically, barcode sequences are encoded at the termini of linear DNA either through ligation or PCR protocols. Following sequencing of pooled samples, each sample can then be demultiplexed, and assigned to their original sample based on these barcode sequences. Several highly multiplexed NGS systems have been developed, permitting the sequencing of hundreds to thousands of samples simultaneously (41). Notably, use of pairwise or asymmetric barcodes, where each terminus receives a specific barcode sequence, is a powerful way of economically processing many samples (42, 43). PCR is often utilized to create these specific combinations, given that primer pairs (encoding barcode sequences) can be combined prior to highly parallel PCR amplification of the target sequences. These primer sequences (encoding both the barcode and template-annealing sequence) must be empirically designed, such that they perform reliably and efficiently with their desired DNA template during the PCR. Consequently, there is a need for standardized workflows with specifically designed barcodes for amplicon sequencing, particularly for the inducible protein expression constructs typically used in synthetic biology.

In this study, new barcoded primers for highly multiplexed nanopore amplicon sequencing were designed for use with the standardized BglBrick expression vectors. DNA assembly in synthetic biology often employs standardized BglBrick expression vectors, each containing a universal MCS (3, 33, 36). Therefore, generic primer binding sequences common to all the BglBrick vectors were selected, such that they were upstream and downstream from the MCS. Consequently, the same sequences could be employed to amplify any insert in any of the BglBrick vectors (Supplementary Material S1). A pairwise (asymmetric) multiplexing approach was employed, in which different 5′ and 3′ barcodes (encoded by the forward and reverse primers, respectively) identify the original sample location, whereby the 5′ barcode identifies the plate origin (termed 'plate barcode' primer) and the 3′ barcode identifies the well in that plate (termed 'well barcode' primer, Figure 1). Of all, 576 bespoke designed primer pairs (a total of six 96-well plates) were generated (see Supplementary Files); a number which can be increased with the introduction of further forward primers (Supplementary Material S3).

PCR amplicon preparation and sequencing

Applying PCR directly from culture is the optimal method to isolate the barcoded construct sequences from a bacterial host, given that the alternative protocol of plasmid extraction, digestion/fragmentation and then barcoding is significantly slower, more labor intensive and costly. To this end, host transformant colonies were selected and cultured overnight, then prepared as PCR templates before addition to PCR reaction mixtures.

The high-fidelity polymerase CloneAmp HiFi (Takara Clonetech) was optimized to establish robust PCR conditions for amplifying constructs up to 10 kb. Every primer pair was tested using an exemplar 6.6 kb construct (Supplementary File) and provided efficient amplification of the target sequence, yielding a single PCR product of the correct size with high yield (Supplementary Material S4). Additionally, this amplicon length analysis provided an early screen for correct assemblies before sequencing, given that amplicons could be compared to their expected pathway lengths.

Having obtained barcoded amplicons, these are pooled and purified, then prepared for nanopore sequencing. This procedure involves end preparation, adapter ligation and incubation

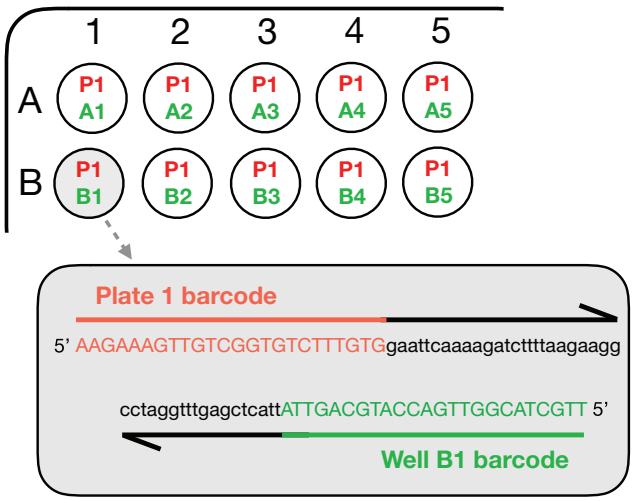

Figure 1. Allocation of primer pairs to enable the identification of individual wells from highly multiplexed samples, using well B1 from plate 1 as an example. Each well is allocated a forward primer, identifying the source plate, and a reverse primer, identifying the well. This enables the accurate identification of each individual well by data analysis after sequencing.

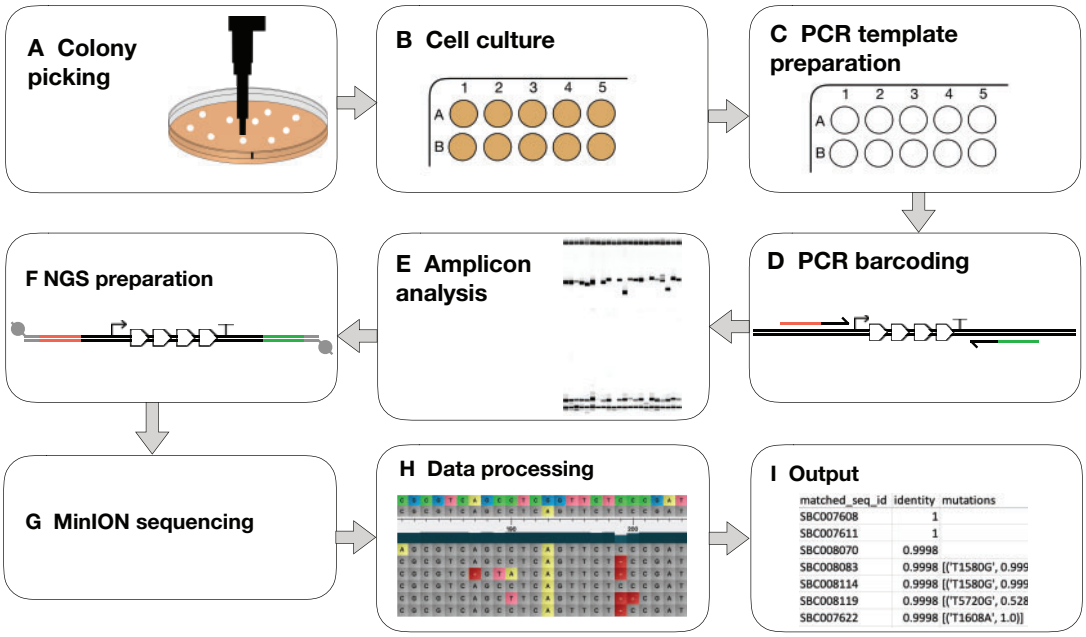

Figure 2. Overview of the construct-sequencing workflow. Colonies harbouring assembled plasmids are first (**A**) picked and (**B**) cultured in deep well plates, prior to (**C**) dilution to create the PCR template. (**D**) PCR amplification of the construct generates 5′ (red) and 3′ (green) barcoded amplicons which are (**E**) analyzed by capillary electrophoresis. (**F**) Pooled amplicons are prepared for NGS sequencing by adapter ligation and (**G**) sequenced using the MinION device. (**H**) Bioinformatics processing of data identifies mutations and removes systematic errors by probabilistic analysis and (**I**) data metrics are outputted.

with the motor protein, a process that prepares the DNA for translocation through the nanopore during the sequencing run (Figure 2).

The data processing algorithm for accurate sequencing using nanopore data

A major limitation of nanopore sequencing data is its low single-passage strand sequencing accuracy, manifesting in a high-error frequency (44, 45). Simple alignment of these erroneous 'noisy' sequences generates a consensus sequence, which given sufficient read depth is capable of eliminating almost all of the sequencing error. However, systematic errors in the data are difficult to remove from standard consensus alignments (46), restricting these data for use in accurate SNV detection. Indeed, in this study some constructs known to be correct (tested by Sanger sequencing) were identified as having only 99.5–99.9% identity to the target sequence from consensus alignment of all sequencing reads, with one or more systematic sequencing errors preventing complete identity.

It is believed that systematic errors can arise from either specific sequences with indistinguishable conductance signals or motor enzyme processing errors (45). As both these events could be sequence-specific, we hypothesized that the

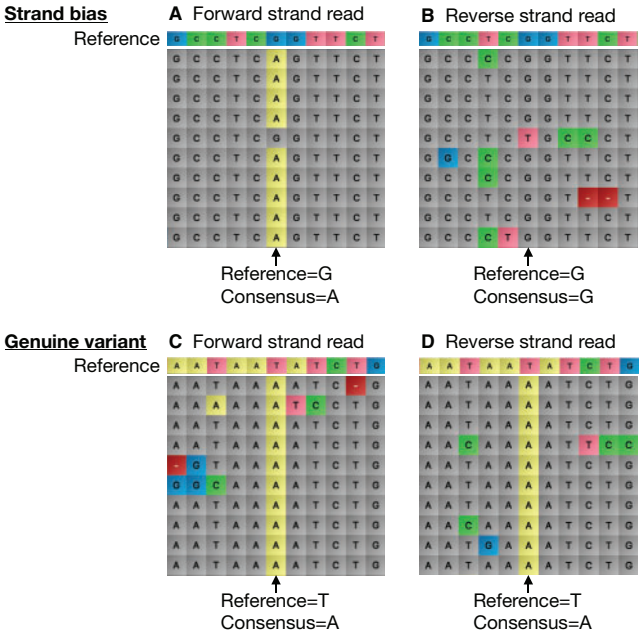

Figure 3. Examples of strand bias and a genuine SNV in nanopore data. Each row corresponds to a different read. Bases correctly aligned to the target sequence are shown in gray and potential mutations are highlighted in color (A = yellow, G = blue, T = pink, C = green, deletion = red). The consensus basecall is identified at the relevant position. Strand bias is shown by the inconsistent SNV basecalling between the (**A**) forward and (**B**) reverse strand reads from the same sample. Our statistical analysis of this alignment prevents erroneous SNV identification. In contrast, genuine SNVs are identified by an agreement between the (**C**) forward and (**D**) reverse strand read data.

occurrence of these errors could vary between the strands being passaged, given that the lower strand is passaged as the reverse complement sequence of the upper strand. For example, if a miscalling occurs at the end of a hairpin in a top strand read, the bottom strand read would correctly basecall this sequence before the hairpin is encountered. This type of strand bias is known to occur for various sequencing platforms including Illumina and Nanopore (47, 48); however, to date it has only been exploited for improving SNV accuracy by McElroy *et al.* (49) using data derived from Roche-454 technology.

Strand bias in nanopore data can be identified through separate alignment of forward and reverse strand reads (Figure 3). Where bias occurs, an SNV is identified in one strand with high frequency whilst the corresponding strand read identifies the non-variant base as the most frequent. This strand bias is exploited in our Bayesian statistical analysis to accurately distinguish between true SNVs and systematic sequencing errors through a probabilistic comparison of forward and reverse strand reads (see Materials and methods section). This therefore enables the analysis to identify when an SNV identified is true (i.e. occurs in both top and bottom strand reads) or miscalled (occurs in one but not both reads). To our knowledge, this is the first example of analyzing and exploiting strand bias in a nanopore analysis workflow for accurate SNV identification.

Experimental verification of the multiplexed workflow

To verify the robustness of our workflow, each of the 576 primer pairs was validated using a control construct (a pinocembrin-producing metabolic pathway (3)). Under experimental conditions, each of the 576 primer pairs demonstrated robust amplification of the target construct with high yield by long-range PCR using the diluted bacterial culture directly as a template (see above and Supplementary Material S4). Analysis showed that each of the 576 samples was identified with 100% coverage and no reported misreads or mutations (identity = 1). It was encouraging to observe that the lowest read depth in this dataset was 46 (total of both forward and reverse strands) yet was still reported as correct, demonstrating that this workflow can accurately identify correct assemblies from a low number of consensus reads (Supplementary Materials S5 and S8).

Demonstration of our workflow for experimental sample sequencing was performed on a library of 563 assembled constructs. Analysis of PCR amplifications by capillary electrophoresis (Supplementary Material S6) showed variation in amplicon size, indicating that some constructs were reliably assembled by LCR, whilst others were not correctly assembled (one or more DNA parts missing). This inconsistency is caused by LCR assembly efficiency, which can be significantly hampered by the number, size, sequence and complexity of the DNA parts being assembled. Upon sequencing, a total of 128 from 563 test assemblies were verified as fully correct sequences with 100% nucleotide identity (Supplementary Materials S7 and S8), in agreement with the size approximations from the electrophoresis analysis (Supplementary Material S6).

Concluding remarks

In this study, a standardized protocol for highly multiplexed DNA sequencing of assembled constructs for synthetic biology is presented. Using PCR amplification directly from bacterial cultures, barcoded amplicons are generated that are ready for sequencing, enabling the rapid processing of hundreds of samples in parallel. Currently, the process is capable of processing data for 576 samples in 72 h (this can be shorter when preparing fewer samples). This process is therefore not only quicker and higher throughput than conventional Sanger sequencing, but cost per sample is also significantly lower. In the workflow, the total cost per sample for the entire workflow is £2.20 ($2.73) regardless of construct length. For the control experiment (6.6 kb

amplicon per sample), the price per kb was £0.33 ($0.41), which is substantially cheaper than the equivalent length Sanger sequencing (currently £7.77 ($9.64) per kb, Supplementary Material S9). Taken together, with the hardware acquisition of the nanopore (the device is provided as part of a 'starter pack'), this presents an attractive low-cost means to perform NGS in-house for small- to medium-sized laboratories.

In addition to cost, this workflow provides other novel features. First, an optimized set of barcoding primers are provided, designed such that they can be repeatedly used with any of the 96 expression plasmids from the BglBrick library (33). While our example study demonstrates the workflow using this library, the same primer design approach is equally applicable to other sets of expression plasmids (by transferring the barcode sequences to any plasmid-specific primer sequence). Furthermore, the design is amenable to processing more samples by adding additional primers sets (further sets for up to 12 × 96 well plates are described in Supplementary Material S3). Aside from the pathway sequencing shown here, this workflow can also be easily applied to other sequencing applications, such as variant libraries in directed evolution and single nucleotide polymorphism analysis and identification of correct sequences during gene synthesis.

Additionally, a powerful informatics algorithm is provided for automated processing of FASTQ files. Typically, a lack of relevant bioinformatics expertise to process NGS data often prevents laboratories from utilizing this powerful technology in-house, and therefore this algorithm provides a useful means for them to exploit this technology. In the analysis of the control dataset, it was found that a read depth of as little as 46 reads was sufficient for accurate sequence verification. Given that the total number of reads (with both barcodes identified) obtained in 24 h were >1 329 000, a theoretical sequencing capacity is calculated of over 10 000 constructs. However, in order to achieve this throughput further effort is required to normalize the sample concentrations to ensure sufficient reads are obtained for every barcode set. Given the continuing demand for higher throughput DNA assembly capability in synthetic biology, the approach introduced here provides the community with a powerful resource for fast multiplexed DNA sequencing and analysis at a dramatically lower cost than for other sequencing technologies.

Availability

All NGS data and processed results described in this work are freely available at https://console.cloud.google.com/storage/browser/sbc-ngs/. Source code, the SBC003382 plasmid sequence and instructions for use are openly available at https://github.com/neilswainston/sbc-ngs/.

Author contributions

A.C., N.S., M.D., A.J.J., C.J.R., S.T., P.C., K.A.H. and C.Y. conceived the project, experimental design and performed experimental work. N.S. and R.B. wrote the analysis algorithm. P.M. established the data management. A.C., N.S., E.T., N.S.S. and R.B. compiled and edited the manuscript and supervised the project.

Supplementary data

Supplementary Data are available at SYNBIO Online.

Acknowledgments

The authors are grateful to Prof. Douglas Kell for initiating the idea of using nanopore sequencing in SYNBIOCHEM and Dr Daniel Schindler for useful discussions.

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) through the grant Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM) [BB/M017702/1]. This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreements 720793 TOPCAPI–Thoroughly Optimised Production Chassis for Advanced Pharmaceutical Ingredients and 814408 SHIKI FACTORY100–Modular cell factories for the production of 100 compounds from the shikimate pathway.

Conflict of interest statement. None declared.

References

1. Paddon,C.J., Westfall,P.J., Pitera,D.J., Benjamin,K., Fisher,K., McPhee,D., Leavell,M.D., Tai,A., Main,A. and Eng,D. (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496, 528–532.
2. Carbonell,P., Currin,A., Jervis,A.J., Rattray,N.J.W., Swainston,N., Yan,C., Takano,E. and Breitling,R. (2016) Bioinformatics for the synthetic biology of natural products: integrating across the Design–Build–Test cycle. *Nat. Prod. Rep.*, 33, 925–932.
3. Carbonell,P., Jervis,A.J., Robinson,C.J., Yan,C., Dunstan,M., Swainston,N., Vinaixa,M., Hollywood,K.A., Currin,A., Rattray,N.J.W. et al. (2018) An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Commun. Biol.*, 1, 66.
4. Khara,B., Menon,N., Levy,C., Mansell,D., Das,D., Marsh,E.N.G., Leys,D. and Scrutton,N.S. (2013) Production of propane and other short-chain alkanes by structure-based engineering of ligand specificity in aldehyde-deformylating oxygenase. *Chembiochem*, 14, 1204–1208.
5. Jang,Y.-S., Park,J.M., Choi,S., Choi,Y.J., Seung,D.Y., Cho,J.H. and Lee,S.Y. (2012) Engineering of microorganisms for the production of biofuels and perspectives based on systems metabolic engineering approaches. *Biotechnol. Adv.*, 30, 989–1000.
6. Ahmed,S.T., Leferink,N.G.H. and Scrutton,N.S. (2019) Chemo-enzymatic routes towards the synthesis of bio-based monomers and polymers. *Mol. Catal.*, 467, 95–110.
7. Roberts,A.D., Finnigan,W., Wolde-Michael,E., Kelly,P., Blaker,J.J., Hay,S., Breitling,R., Takano,E. and Scrutton,N.S. (2019) Synthetic biology for fibers, adhesives, and active camouflage materials in protection and aerospace. *MRS Commun.*, 9, 486–504.
8. Trabelsi,H., Koch,M. and Faulon,J.-L. (2018) Building a minimal and generalizable model of transcription-factor based biosensors: showcasing flavonoids. *Biotechnol. Bioeng*, 115, 2292–2304.
9. Currin,A., Korovin,K., Ababi,M., Roper,K., Kell,D.B., Day,P.J. and King,R.D. (2017) Computing exponentially faster: implementing a non-deterministic universal Turing machine using DNA. *J. R. Soc. Interface*, 14, 20160990.
10. Ellis,T., Adie,T. and Baldwin,G.S. (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol.*, 3, 109–118.

11. Casini,A., Storch,M., Baldwin,G.S. and Ellis,T. (2015) Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.*, 16, 568–576.
12. Cobb,R.E., Ning,J.C. and Zhao,H. (2014) DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J. Ind. Microbiol. Biotechnol.*, 41, 469–477.
13. Suckling,L., McFarlane,C., Sawyer,C., Chambers,S.P., Kitney,R.I., McClymont,D.W. and Freemont,P.S. (2019) Miniaturisation of high-throughput plasmid DNA library preparation for next-generation sequencing using multifactorial optimisation. *Synth. Syst. Biotechnol.*, 4, 57–66.
14. D'Amore,R., Johnson,J., Haldenby,S., Hall,N., Hughes,M., Joynson,R., Kenny,J.G., Patron,N., Hertz-Fowler,C. and Hall,A. (2017) SMRT Gate: a method for validation of synthetic constructs on Pacific Biosciences sequencing platforms. *BioTechniques*, 63, 13–20.
15. Lucks,J.B., Mortimer,S.A., Trapnell,C., Luo,S., Aviran,S., Schroth,G.P., Pachter,L., Doudna,J.A. and Arkin,A.P. (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. USA*, 108, 11063–11068.
16. Takahashi,M.K., Watters,K.E., Gasper,P.M., Abbott,T.R., Carlson,P.D., Chen,A.A. and Lucks,J.B. (2016) Using in-cell SHAPE-Seq and simulations to probe structure-function design principles of RNA transcriptional regulators. *RNA*, 22, 920–933.
17. Watters,K.E., Abbott,T.R. and Lucks,J.B. (2016) Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic Acids Res.*, 44, e12.
18. Gorochowski,T.E., Chelysheva,I., Eriksen,M., Nair,P., Pedersen,S. and Ignatova,Z. (2019) Absolute quantification of translational regulation and burden using combined sequencing approaches. *Mol. Syst. Biol.*, 15, e8719.
19. Fernandez-Rodriguez,J., Yang,L., Gorochowski,T.E., Gordon,D.B. and Voigt,C.A. (2015) Memory and combinatorial logic based on DNA inversions: dynamics and evolutionary stability. *ACS Synth. Biol.*, 4, 1361–1372.
20. Liu,Q., Schumacher,J., Wan,X., Lou,C. and Wang,B. (2018) Orthogonality and burdens of heterologous AND gate gene circuits in *E. coli*. *ACS Synth. Biol.*, 7, 553–564.
21. Gorochowski,T.E., Espah Borujeni,A., Park,Y., Nielsen,A.A., Zhang,J., Der,B.S., Gordon,D.B. and Voigt,C.A. (2017) Genetic circuit characterization and debugging using RNA-seq. *Mol. Syst. Biol*, 13, 952.
22. Ansorge,W.J. (2009) Next-generation DNA sequencing techniques. *New Biotechnol.*, 25, 195–203.
23. Liu,L., Li,Y., Li,S., Hu,N., He,Y., Pong,R., Lin,D., Lu,L. and Law,M. (2012) Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.*, 2012, 1–11.
24. Ashton,P.M., Nair,S., Dallman,T., Rubino,S., Rabsch,W., Mwaigwisya,S., Wain,J. and O'Grady,J. (2015) MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat. Biotech.*, 33, 296–300.
25. Xu,P., Rizzoni,E.A., Sul,S.-Y. and Stephanopoulos,G. (2017) Improving metabolic pathway efficiency by statistical model-based multivariate regulatory metabolic engineering. *ACS Synth. Biol.*, 6, 148–158.
26. Jeschek,M., Gerngross,D. and Panke,S. (2016) Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nat. Commun.*, 7, 11163.
27. Rang,F.J., Kloosterman,W.P. and de Ridder,J. (2018) From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.*, 19, 90.
28. Goodwin,S., Gurtowski,J., Ethe-Sayers,S., Deshpande,P., Schatz,M.C. and McCombie,W.R. (2015) Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.*, 25, 1750–1756.
29. Ma,Z.S., Li,L., Ye,C., Peng,M. and Zhang,Y.P. (2018) Hybrid assembly of ultra-long Nanopore reads augmented with 10x-Genomics contigs: Demonstrated with a human genome. *Genomics*, pii: S0888-7543(18)30560-3. doi: 10.1016/j.ygeno.2018.12.013.
30. De Maio,N., Shaw,L.P., Hubbard,A., George,S., Sanderson,N.D., Swann,J., Wick,R., AbuOun,M., Stubberfield,E. and Hoosdally,S.J. et al (2019) Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb Genom.*, 5.
31. Morisse,P., Lecroq,T. and Lefebvre,A. (2018) Hybrid correction of highly noisy long reads using a variable-order de Bruijn graph. *Bioinformatics*, 34, 4213–4222.
32. Vasudevan,K., Devanga Ragupathi,N.K., Jacob,J.J., Veeraraghavan,B. (2019) Highly accurate-single chromosomal complete genomes using IonTorrent and MinION sequencing of clinical pathogens. *Genomics*. pii: S0888-7543(19)30058-8.
33. Lee,T.S., Krupa,R.A., Zhang,F., Hajimorad,M., Holtz,W.J., Prasad,N., Lee,S.K. and Keasling,J.D. (2011) BglBrick vectors and datasheets: a synthetic biology platform for gene expression. *J. Biol. Eng.*, 5, 12.
34. Markham,N.R. and Zuker,M. (2005) DINAMelt web server for nucleic acid melting prediction. *Nucl. Acids Res.*, 33, W577–W581.
35. Jervis,A.J., Carbonell,P., Vinaixa,M., Dunstan,M.S., Hollywood,K.A., Robinson,C.J., Rattray,N.J.W., Yan,C., Swainston,N., Currin,A. et al. (2019) Machine learning of designed translational control allows predictive pathway optimization in *Escherichia coli*. *ACS Synth. Biol.*, 8, 127–136.
36. Robinson,C.J., Dunstan,M.S., Swainston,N., Titchmarsh,J., Takano,E., Scrutton,N.S. and Jervis,A.J. (2018) Chapter Thirteen: Multifragment DNA assembly of biochemical pathways via automated ligase cycling reaction. In: N. Scrutton (ed). *Methods in Enzymology, Enzymes in Synthetic Biology*, Vol. 608. Academic Press, Cambridge, Massachusetts, pp. 369–392.
37. Loose,M., Malla,S. and Stout,M. (2016) Real-time selective sequencing using nanopore technology. *Nat. Methods*, 13, 751–754.
38. Cock,P.J.A., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B. et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25, 1422–1423.
39. Li,H. and Durbin,R. (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 26, 589–595.
40. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.
41. Smith,A.M., Heisler,L.E., St.Onge,R.P., Farias-Hesson,E., Wallace,I.M., Bodeau,J., Harris,A.N., Perry,K.M., Giaever,G., Pourmand,N. et al. (2010) Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Res.*, 38, e142.
42. Tu,J., Ge,Q., Wang,S., Wang,L., Sun,B., Yang,Q., Bai,Y. and Lu,Z. (2012) Pair-barcode high-throughput sequencing for

large-scale multiplexed sample analysis. *BMC Genomics*, 13, 43.
43. Srivathsan,A., Baloğlu,B., Wang,W., Tan,W.X., Bertrand,D., Ng,A.H.Q., Boey,E.J.H., Koh,J.J.Y., Nagarajan,N. and Meier,R. (2018) A MinION™-based pipeline for fast and cost-effective DNA barcoding. *Mol. Ecol. Res.*, 18, 1035–1049.
44. Branton,D., Deamer,D.W., Marziali,A., Bayley,H., Benner,S.A., Butler,T., Di Ventra,M., Garaj,S., Hibbs,A., Huang,X. et al. (2008) The potential and challenges of nanopore sequencing. *Nat. Biotechnol.*, 26, 1146–1153.
45. Noakes,M.T., Brinkerhoff,H., Laszlo,A.H., Derrington,I.M., Langford,K.W., Mount,J.W., Bowman,J.L., Baker,K.S., Doering,K.M., Tickman,B.I. et al. (2019) Increasing the accuracy of nanopore DNA sequencing using a time-varying cross membrane voltage. *Nat. Biotechnol.*, 37, 651–656.
46. Krishnakumar,R., Sinha,A., Bird,S.W., Jayamohan,H., Edwards,H.S., Schoeniger,J.S., Patel,K.D., Branda,S.S. and Bartsch,M.S. (2018) Systematic and stochastic influences on the performance of the MinION nanopore sequencer across a range of nucleotide bias. *Sci. Rep.*, 8, 3159.
47. Guo,Y., Li,J., Li,C.-I., Long,J., Samuels,D.C. and Shyr,Y. (2012) The effect of strand bias in Illumina short-read sequencing data. *BMC Genomics*, 13, 666.
48. Xu,C. (2018) A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Comput. Struct. Biotechnol. J.*, 16, 15–24.
49. McElroy,K., Zagordi,O., Bull,R., Luciani,F. and Beerenwinkel,N. (2013) Accurate single nucleotide variant detection in viral populations by combining probabilistic clustering with a statistical test of strand bias. *BMC Genomics*, 14, 501.